

Routine versus clinically driven laboratory monitoring and first-line antiretroviral therapy strategies in African children with HIV (ARROW): a 5-year open-label randomised factorial trial



ARROW Trial team*

Summary

Background No trials have investigated routine laboratory monitoring for children with HIV, nor four-drug induction strategies to increase durability of first-line antiretroviral therapy (ART).

Methods In this open-label parallel-group trial, Ugandan and Zimbabwean children or adolescents with HIV, aged 3 months to 17 years and eligible for ART, were randomly assigned in a factorial design. Randomisation was to either clinically driven monitoring or routine laboratory and clinical monitoring for toxicity (haematology and biochemistry) and efficacy (CD4 cell counts; non-inferiority monitoring randomisation); and simultaneously to standard three-drug or to four-drug induction first-line ART, in three groups: three-drug treatment (non-nucleoside reverse transcriptase inhibitor [NNRTI], lamivudine, abacavir; group A) versus four-drug induction (NNRTI, lamivudine, abacavir, zidovudine; groups B and C), decreasing after week 36 to three-drug NNRTI, lamivudine, plus abacavir (group B) or lamivudine, abacavir, plus zidovudine (group C; superiority ART-strategy randomisation). For patients assigned to routine laboratory monitoring, results were returned every 12 weeks to clinicians; for clinically driven monitoring, toxicity results were only returned for requested clinical reasons or if grade 4. Children switched to second-line ART for WHO stage 3 or 4 events or (routine laboratory monitoring only) age-dependent WHO CD4 criteria. Randomisation used computer-generated sequentially numbered tables incorporated securely within the database. Primary efficacy endpoints were new WHO stage 4 events or death for monitoring and change in CD4 percentage at 72 and 144 weeks for ART-strategy randomisations; the co-primary toxicity endpoint was grade 3 or 4 adverse events. Analysis was by intention to treat. This trial is registered, ISRCTN24791884.

Findings 1206 children were randomly assigned to clinically driven ($n=606$) versus routine laboratory monitoring ($n=600$), and groups A ($n=397$), B ($n=404$), and C ($n=405$). 47 (8%) children on clinically driven monitoring versus 39 (7%) on routine laboratory monitoring had a new WHO stage 4 event or died (hazard ratio [HR] 1.13, 95% CI 0.73–1.73, $p=0.59$; non-inferiority criterion met). However, in years 2–5, rates were higher in children on clinically driven monitoring (1.3 vs 0.4 per 100 child-years, difference 0.99, 0.37–1.60, $p=0.002$). One or more grade 3 or 4 adverse events occurred in 283 (47%) children on clinically driven versus 282 (47%) on routine laboratory monitoring (HR 0.98, 0.83–1.16, $p=0.83$). Mean CD4 percentage change did not differ between ART groups at week 72 (16.5% [SD 8.6] vs 17.1% [8.5] vs 17.3% [8.0], $p=0.33$) or week 144 ($p=0.69$), but four-drug groups (B, C) were superior to three-drug group A at week 36 (12.4% [7.2] vs 14.1% [7.1] vs 14.6% [7.3], $p<0.0001$). Excess grade 3 or 4 events in groups B (one or more events reported by 157 [40%] children in A, 190 [47%] in B; HR [B:A] 1.32, 1.07–1.63) and C (218 [54%] children in C; HR [C:A] 1.58, 1.29–1.94; global $p=0.0001$) were driven by asymptomatic neutropenia in zidovudine-containing groups (B, C; 86 group A, 133 group B, 184 group C), but resulted in drug substitutions in only zero versus two versus four children, respectively.

Interpretation NNRTI plus NRTI-based three-drug or four-drug ART can be given across childhood without routine toxicity monitoring; CD4 monitoring provided clinical benefit after the first year on ART, but event rates were very low and long-term survival high, suggesting ART rollout should take priority. CD4 benefits from four-drug induction were not durable, but three-NRTI long-term maintenance was immunologically and clinically similar to NNRTI-based ART and could be valuable during tuberculosis co-treatment.

Funding UK Medical Research Council, the UK Department for International Development; drugs donated and viral load assays funded by ViiV Healthcare and GlaxoSmithKline.

Introduction

In high-income countries, routine laboratory tests are done (typically every 3–4 months) to monitor efficacy (HIV viral load, CD4 counts) and toxicity (haematology

or biochemistry panels) in patients with HIV on antiretroviral therapy (ART). Although not mandated when public health ART rollout started,¹ the extent to which HIV treatment programmes in resource-limited

Lancet 2013; 381: 1391–403

Published Online

March 7, 2013

[http://dx.doi.org/10.1016/S0140-6736\(12\)62198-9](http://dx.doi.org/10.1016/S0140-6736(12)62198-9)

See [Comment](#) page 1343

See [Online](#) for a video about the ARROW clinical trial

*See appendix for Analysis and Writing Committee

Correspondence to:

Prof Diana M Gibb, MRC Clinical Trials Unit, London WC2B 6NH, UK
dmg@ctu.mrc.ac.uk

See [Online](#) for appendix

settings should provide routine laboratory monitoring versus focus resources on expanding ART access for the many in need, remains an ongoing debate, particularly relevant with slowing growth in health assistance for resource-limited countries in the current economic crisis.²

90% of children with HIV live in sub-Saharan Africa; in 2011, only about 28% of those needing ART were receiving it.³ Infrastructure, personnel, and supply chain constraints all affect the ability to monitor, and tests for detecting toxicity and measuring efficacy are generally unavailable, particularly at low-level facilities.⁴ Trials in adults have compared routine laboratory with clinical monitoring on ART.^{5–8} These showed no benefit from routine toxicity monitoring,⁵ small but significant benefits from CD4 monitoring,^{5,8} and no significant additional benefit of viral load over CD4 monitoring.^{6,8} No trials have evaluated monitoring strategies in children on ART; results in adults might not be generalisable because of differences in frequency of ART toxicity, predictive value of CD4s,^{9,10} and the clinical spectrum of paediatric HIV and its comorbidities, particularly in Africa.¹¹

Children with HIV start ART early and will need to take treatment for longer than adults. Challenges in treatment of young children, in whom viral loads are higher, pharmacokinetics more variable, and adherence more challenging, might account for the lower virological suppression rates in children compared with adults in similar calendar periods.^{12,13} However, four-drug combinations (generally three nucleoside reverse transcriptase inhibitors [NRTIs] plus one non-nucleoside reverse transcriptase inhibitor [NNRTI]) have been reported to show superior virological and immunological responses compared with three-drug ART (two NRTIs plus one NNRTI or protease inhibitor) in an observational study in infants.¹⁴ Whether addition of one drug for a short period in an induction-maintenance approach might have benefits across childhood is unknown. A key advantage might be more rapid reduction of high viral loads in ART-naïve children, which could have long-term benefits. A triple NRTI regimen after four-drug NRTI plus NNRTI induction also avoids challenges of managing interactions with antituberculosis drugs, particularly for young children unable to take efavirenz.

We therefore undertook the first paediatric trial (AntiRetroviral Research fOr Watoto [ARROW]) to evaluate the long-term effect of routine efficacy and toxicity monitoring and different first-line ART strategies in a factorial design in African children.

Methods

Trial design

ARROW was an open-label randomised parallel-group trial in untreated (except for ART to prevent mother-to-child-transmission) children or adolescents (aged 3 months to 17 years) with HIV who met WHO 2006 criteria for ART initiation¹⁵ from three centres in Uganda

(Joint Clinical Research Centre, Kampala; Baylor-Uganda, Mulago; MRC/UVRI Uganda Research Unit on AIDS, Entebbe), and one in Zimbabwe (University of Zimbabwe, Harare). Children with acute infections, on drugs contraindicated with ART, unlikely to adhere or to attend regularly, with laboratory abnormalities contraindicating ART, pregnant or breastfeeding, or perinatally exposed to ART (if <6 months old) were excluded. Participants were randomly assigned (1:1) to clinically driven monitoring versus routine laboratory plus clinical monitoring for toxicity (haematology and biochemistry) and efficacy (CD4). Children were also randomly assigned (1:1:1) in a factorial design to three approaches for first-line ART: open-label lamivudine, abacavir, plus NNRTI continuously (group A, control); induction-maintenance with four-drug lamivudine, abacavir, NNRTI, plus zidovudine for 36 weeks, then lamivudine, abacavir, plus NNRTI (group B); or induction-maintenance with lamivudine, abacavir, NNRTI, plus zidovudine for 36 weeks, then lamivudine, abacavir, plus zidovudine (group C). HLA testing was not done. HIV viral loads were done retrospectively on stored samples. The hypothesis was that clinically driven monitoring would result in similar outcomes to routine laboratory monitoring (non-inferiority) and that four-drug induction-maintenance would have greater efficacy than would standard three-drug ART. The NNRTI (nevirapine or efavirenz) was chosen by clinicians according to local availability and age. Caregivers gave written informed consent; older children (8–17 years) aware of their HIV status also gave assent or consent following national guidelines. The trial was approved by research ethics committees in Uganda, Zimbabwe, and the UK.

Randomisation and masking

Both factorial randomisations were stratified by centre and age (<7, 7–12, ≥13 years). The computer-generated sequentially numbered randomisation list (with variable block sizes) containing both allocations was prepared by the trial statistician and incorporated within the secure database at each trial centre, connected to but not located within each clinical centre, allowing trial managers to access the next number, but not the whole list. Randomisation was undertaken by clinicians phoning the local trials centre. Once randomised, allocation was open—ie, physicians and carers were aware of group assignment.

Procedures

All participants were examined by a doctor and had routine full blood count with white cell differential, lymphocyte subsets (CD4, CD8), biochemistry tests (bilirubin, urea, creatinine, aspartate aminotransferase, alanine aminotransferase) at screening, randomisation (lymphocytes only), weeks 4, 8, and 12, then every 12 weeks. Screening results were used to assess eligibility. All subsequent results for participants assigned to

routine laboratory monitoring were returned to clinicians, whereas results at and after randomisation for participants allocated clinically driven monitoring were only returned if requested for clinical management (authorised by centre project leaders); haemoglobin results at week 8 were automatically returned on the basis of early anaemia in DART,¹⁶ as were grade 4 laboratory toxicities (protocol safety criteria; grades defined¹⁷ apart from neutrophils¹⁸). Total lymphocytes and CD4 tests were never returned for participants on clinically driven monitoring, but for all children other investigations could be requested and concomitant drugs prescribed, as clinically indicated.

All children received ART as syrups or tablets dosed according to WHO weight-band tables.^{19–21} Children were reviewed every 4–6 weeks by a nurse using a standard symptom checklist. Antiretroviral drugs could be substituted, preferably within class, after adverse events. During four-drug induction, a drug could be dropped because of toxicity or drug interactions with anti-tuberculosis treatment. Switching to second-line ART (including a ritonavir-boosted protease inhibitor) was based on clinical criteria in all participants (new or recurrent WHO stage 4 event;¹⁵ or WHO stage 3 event or events at clinician discretion, particularly if recurrent or persistent), or on laboratory criteria for routine laboratory monitoring (confirmed on-ART CD4 of <15% at age 1–2 years, <10% at 3–4 years, <100 cells/mL at ≥5 years). See appendix for further details.

Co-primary endpoints for the monitoring randomisation were progression to new WHO stage 4 event or death (efficacy), and grade 3 or 4 adverse events not solely related to HIV (safety). Co-primary endpoints for the ART-strategy randomisation were change in CD4 percentage from randomisation to 72 and 144 weeks (efficacy) and grade 3 or 4 adverse events (safety). Secondary endpoints for both randomisations (if not co-primary) were: mortality; new (or new or recurrent) WHO stage 4 event or death; new (or new or recurrent) WHO stage 3 or 4 event or death; grade 3 or 4 adverse events definitely, probably, or uncertainly ART-related; serious adverse events²² not solely HIV-related; ART-modifying adverse events; admissions to hospital; height, weight, and body-mass index for age; CD4; number and class of antiretrovirals received; switch to second-line regimen; adherence; viral load and resistance (done retrospectively; see appendix). All WHO stage 3 or 4 events, deaths, and serious adverse events were reviewed against prespecified criteria by an endpoint review committee with an independent chair and members, masked to randomised allocations.

Statistical analysis

1200 children followed up for 3·5–5·0 years with less than 10% loss to follow-up provided 90% power to establish that clinically driven monitoring was not inferior to routine laboratory monitoring on the primary

efficacy outcome, defined as the upper 95% confidence limit for the difference (clinically driven minus routine laboratory monitoring) in rate of first new WHO stage 4 event or deaths per 100 child-years being no greater than 1·6 per 100 child-years (assumed rate for routine laboratory monitoring of 2·5 per 100 child-years). For the ART-strategy randomisation, 1200 children provided 80% power to detect differences in change in CD4 percentage of more than 2·5% across the three groups (*F* test, two-sided $\alpha=0\cdot05$) assuming 20% missing data (loss to follow-up, missed visit or test) and standard deviation 10%.²³ Interim data were reviewed annually by an independent data monitoring committee (four meetings) using the Haybittle-Peto criterion ($p<0\cdot001$).

Randomised groups were compared with Kaplan-Meier plots, log-rank tests, and proportional hazards models, stratified by randomisation stratification factors (including the other factorial) for time-to-event disease progression, ART, and adverse event outcomes, censoring at the earlier of trial closure or last follow-up. Categorical variables were compared with χ^2 or exact (if indicated) tests. Change in CD4 percentage was compared with normal linear regression adjusted for randomisation stratification factors. Laboratory measurements and adherence were compared across randomised groups over time with generalised estimating equations (independent correlation; closest measurement to each scheduled visit within equally spaced windows). All comparisons were as randomised (intention-to-treat). Subgroups specified in the analysis plan were the factorial randomisations, time on ART, sex, age, centre, CD4, weight for age, randomisation year, and previous ART for prevention of mother-to-child transmission. Baseline values were those nearest to but before and within 42 days of randomisation. *Z* scores were determined with the British reference because it covers the full age range of ARROW children.²⁴ All analyses were done with Stata 12.1. All *p* values are two-sided.

This trial is registered, ISRCTN24791884.

Role of the funding sources

The sponsor (UK Medical Research Council), other funders (UK Department for International Development), and ViiV Healthcare/GlaxoSmithKline (donated drugs; funded viral load assays) had no direct role in study design, data collection, analysis, interpretation, report writing, or the decision to submit for publication. The corresponding author had full access to all data and responsibility for submission for publication.

Results

1210 children were enrolled (March 15, 2007–Nov 18, 2008). One was randomly assigned twice at different centres (second randomisation was excluded), and three had major eligibility violations, leaving 600 children assigned to routine laboratory and 606 to clinically driven monitoring, and 397 to group A, 404 to group B, and

405 to group C (figure 1). Baseline characteristics were similar across randomised groups (table 1). Median follow-up to planned trial closure (March 16, 2012) was 4.0 (IQR 3.7–4.4) years in each monitoring and ART-strategy group (maximum 5.0 years; total

4685 child-years). Only 33 (3%) children (median follow-up 150 [IQR 46–169] weeks) not known to have died were not seen after trial closure (figure 1). Completeness of nurse visits every 4–6 weeks and doctor visits every 12 weeks was more than 95% (46 531/48 461 nurse, 19 088/19 765 doctor visits) and was similar in all groups.

For children assigned to clinically driven monitoring, clinicians could request individual laboratory toxicity results from routine haematology or biochemistry panels for clinical reasons. Panel tests could also be requested in both groups at extra patient-initiated visits. However, most were done at routine visits (10 805 [92%] haematology and 10 778 [94%] biochemistry; appendix). In clinically driven monitoring, apart from week 8 haemoglobin (returned per protocol), very few results were released (from 486 [4%] panels); most commonly requested results were haemoglobin (2%, $n=265$) and neutrophils (3%, $n=323$; appendix). More additional haematology tests were requested during nurse visits or extra visits in routine laboratory than in clinically driven monitoring ($p<0.0001$); there were no significant differences in requests for additional biochemistry tests ($p=0.17$) or other non-routine (eg, electrolytes) tests ($p=0.97$).

At trial end, 578 (95%) children on clinically driven monitoring versus 565 (94%) on routine laboratory monitoring were still on first-line ART (table 2); 330 (83%) children in group A, 356 (88%) in group B, and 367 (91%) in group C were still on their original randomised regimen. Among 151 first-line drug changes, equal numbers were due to adverse events and antituberculosis therapy (both 59 changes [39%]; table 2). Changes to first-line therapy occurred at rates of 3.3 per 100 child-years in routine laboratory monitoring versus 3.2 per 100 child-years in clinically driven monitoring ($p=0.94$); and 3.5, 3.7, and 2.6 per 100 child-years in groups A, B, and C, respectively ($p=0.25$).

30 children in group A, 20 in group B, and nine in group C stopped nevirapine because of starting anti-tuberculosis treatment. Whereas in group A nevirapine was mainly substituted with zidovudine (<3 years) or efavirenz (>3 years), in the four-drug groups, about a third (five children in group B and six in C) simply dropped nevirapine during four-drug induction to continue on three NRTIs. Adverse events resulted in a drug being substituted or dropped in five children in group A, 28 in group B, and 26 in group C; most changes were zidovudine-related (16 in B, 20 in C; table 2).

Adherence by self-reported questionnaire was similar in both monitoring groups: mean 6.7% (1813/26 917) of children on clinically driven monitoring reported missing doses in the past 28 days versus 6.5% (1683/25 935) on routine laboratory monitoring ($p=0.26$). There were small but significant differences by ART strategy, with fewer reporting missing doses in the past 28 days in group A through week 36 (induction period; 7.9% [253/2966] in group A, 9.8% [326/3014] in group B, 9.0% [295/2999] in group C; $p=0.02$) and overall

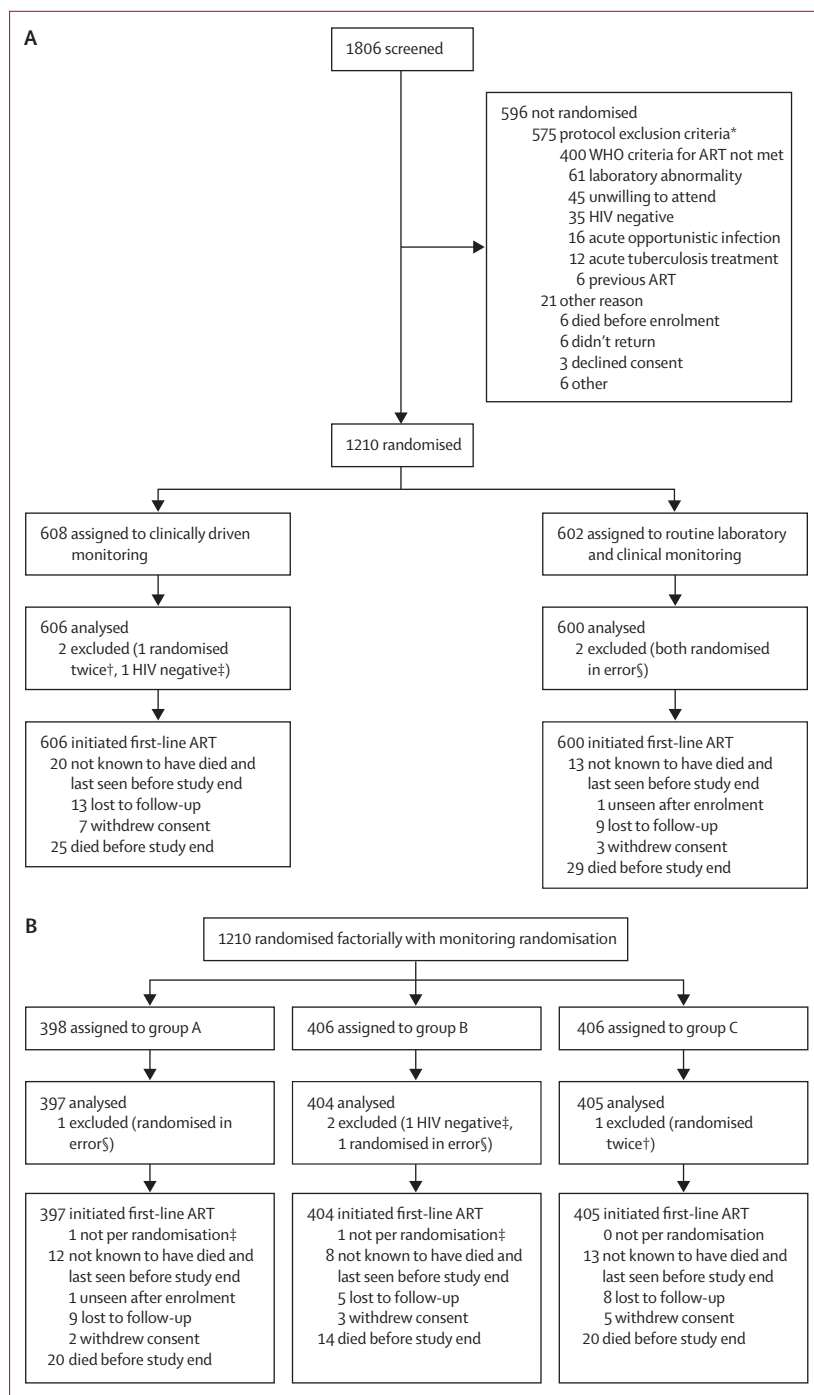


Figure 1: Trial profile

Study ended on March 16, 2012. ART=antiretroviral therapy. *Predominant reason for ineligibility given according to frequency. †First randomisation included, second excluded. ‡Communication error at enrolment. §Both on intensive phase of tuberculosis treatment.

(6·1% [1059/16345], 7·3% [1302/16640], 6·5% [1135/16371], respectively; $p < 0·0001$).

No child switched to second-line treatment during their first year on ART. Overall, 28 (5%) children on clinically driven monitoring versus 35 (6%) on routine laboratory monitoring switched to second-line treatment (hazard ratio [HR] 0·78, 95% CI 0·48–1·29, $p = 0·22$; table 2; figure 2), after median 2·8 (IQR 1·8–3·3) years versus 2·2 (1·5–3·0) years, respectively. All children on

routine laboratory monitoring meeting CD4 switch criteria actually switched. At switch, median CD4 percentage was 8·5% (IQR 1·5–22·5) in the clinically driven monitoring group and 7% (3–13) in the routine laboratory monitoring group; 11 children (39%) versus 14 (40%) had CD4 percentage less than 5% and six children (21%) versus none had CD4 percentage greater than 25%, respectively. 2% of follow-up (40/2373 child-years) was spent on second-line treatment

	Monitoring strategy		First-line ART strategy		
	Clinically driven monitoring (n=606)	Routine laboratory and clinical monitoring (n=600)	Group A (n=397)	Group B (n=404)	Group C (n=405)
Centre					
Entebbe, Uganda	92 (15%)	96 (16%)	61 (15%)	61 (15%)	66 (16%)
Joint Clinical Research Centre, Kampala, Uganda	161 (27%)	157 (26%)	104 (26%)	108 (27%)	106 (26%)
Paediatric Infectious Disease Centre, Mulago, Uganda	152 (25%)	148 (25%)	101 (25%)	99 (25%)	100 (25%)
University of Zimbabwe, Harare, Zimbabwe	201 (33%)	199 (33%)	131 (33%)	136 (34%)	133 (33%)
Male	298 (49%)	298 (50%)	193 (49%)	207 (51%)	196 (48%)
Age (years)	5·9 (2·2 to 9·2)	6·0 (2·6 to 9·4)	6·1 (2·4 to 9·3)	6·2 (2·5 to 9·4)	5·7 (2·3 to 9·3)
Children younger than 3 years	197 (33%)	173 (29%)	121 (30%)	117 (29%)	132 (33%)
Vertical exposure*	605 (>99%)	594 (99%)	395 (99%)	401 (99%)	403 (>99%)
CD4					
Median CD4 percentage (%)	12·5% (7·5 to 17·3)	12·0% (7·0 to 17·0)	11·7% (6·9 to 17·5)	12·0% (7·1 to 17·0)	12·5% (8·0 to 17·0)
Children with CD4 percentage <5%	92 (15%)	110 (18%)	75 (19%)	67 (17%)	60 (15%)
Median CD4 cell count†	244 (114 to 369)	237 (94 to 366)	221 (87 to 341)	242 (108 to 369)	262 (114 to 373)
WHO stage					
1	13 (2%)	7 (1%)	4 (1%)	7 (2%)	9 (2%)
2	160 (26%)	174 (29%)	119 (30%)	116 (29%)	99 (24%)
3	346 (57%)	337 (56%)	217 (55%)	228 (56%)	238 (59%)
4	87 (14%)	82 (14%)	57 (14%)	53 (13%)	59 (15%)
Eligibility for ART					
Clinical only	95 (16%)	96 (16%)	67 (17%)	59 (15%)	65 (16%)
Immunological only	160 (26%)	167 (28%)	114 (29%)	114 (28%)	99 (24%)
Clinical and immunological	351 (58%)	337 (56%)	216 (54%)	231 (57%)	241 (60%)
Median weight-for-age Z score‡	-2·3 (-3·4 to -1·3)	-2·2 (-3·3 to -1·3)	-2·3 (-3·5 to -1·3)	-2·2 (-3·2 to -1·4)	-2·2 (-3·4 to -1·3)
Weight-for-age Z score <-3	191 (32%)	184 (31%)	130 (33%)	119 (29%)	126 (31%)
Median height-for-age Z score‡	-2·5 (-3·4 to -1·5)	-2·4 (-3·4 to -1·5)	-2·4 (-3·4 to -1·5)	-2·5 (-3·4 to -1·5)	-2·4 (-3·3 to -1·5)
Height-for-age Z score <-3	211 (35%)	191 (32%)	139 (35%)	127 (31%)	136 (34%)
Haemoglobin (g/L)	105 (95 to 115)	106 (96 to 116)	105 (94 to 115)	107 (96 to 116)	106 (94 to 115)
Neutrophils (×1000 cells per mL)	2·0 (1·4 to 2·9)	2·1 (1·5 to 2·9)	2·0 (1·4 to 2·8)	2·2 (1·5 to 3·0)	2·0 (1·4 to 2·8)
First-line ART, abacavir plus lamivudine plus					
Nevirapine	129 (21%)	125 (21%)	253 (64%)	1 (<1%)	0
Efavirenz§	71 (12%)	72 (12%)	143 (36%)	0	0
Zidovudine plus nevirapine¶	256 (42%)	248 (41%)	1 (<1%)	249 (62%)	254 (63%)
Zidovudine plus efavirenz¶	150 (25%)	155 (26%)	0	154 (38%)	151 (37%)
Prophylaxis use from enrolment					
Co-trimoxazole	603 (>99%)	598 (>99%)	396 (>99%)	402 (>99%)	403 (>99%)
Dapsone	3 (<1%)	0	1 (<1%)	1 (<1%)	1 (<1%)
None	0	2 (<1%)	0	1 (<1%)	1 (<1%)

Data are n (%) or median (IQR). ART=antiretroviral therapy. *Seven children with missing route of HIV transmission had no evidence of HIV infection in the biological mother who was already deceased at the time of trial enrolment with no other evidence suggesting infection. †Age older than 5 years at enrolment. ‡Z score using UK 1990 reference²⁴ to include full age range. §Efavirenz not licensed for children younger than 3 years. ¶Children initiating ART with four drugs dropped either nevirapine, efavirenz, or zidovudine at week 36 in factorial randomisation. ||Grade 2 neutropenia at enrolment: initiated co-trimoxazole 1 and 4 months later.

Table 1: Characteristics at randomisation

in children on clinically driven monitoring versus 3% (67/2311 child-years) in those on routine laboratory monitoring. 26 [7%] children in group A switched to second-line treatment versus 17 [4%] in group B and 20 (5%) in group C; these differences were not significant ($p=0.25$; figure 2).

There was no evidence of interaction between monitoring and induction-maintenance ART strategies in primary or secondary outcomes (heterogeneity $p>0.1$) except for WHO stage 4 events or death in the first year on ART, for which the relative difference between clinically driven and routine laboratory monitoring varied by first-line ART strategy (appendix). Since most variation was between groups B and C, who received identical ART for the first 36 weeks, this year 1 interaction appeared attributable to chance in children initiating ART with severe immunodeficiency.

47 (8%) children on clinically driven monitoring versus 39 (7%) on routine laboratory monitoring had a new WHO stage 4 event or died (2.0 vs 1.7 per 100 child-years, respectively). The absolute difference of 0.32 per 100 child-years (95% CI -0.47 to 1.12) translated into a relative HR of 1.13 (95% CI 0.73–1.73, $p=0.59$; figure 3). The upper 95% confidence limit for the absolute difference was below the non-inferiority margin of 1.6.

Although overall progression was similar, a prespecified subgroup analysis showed children on routine laboratory monitoring had higher event rates during the first 3 months on ART and lower rates from the second year on ART compared with those on clinically driven monitoring (heterogeneity $p=0.045$; appendix). After year 1, 23 (4%) children on clinically driven monitoring versus six (1%) on routine laboratory monitoring had a first WHO stage 4 event or died (difference 0.99 per

	Monitoring strategy		First-line ART strategy		
	Clinically driven monitoring (n=606)	Routine laboratory and clinical monitoring (n=600)	Group A (n=397)	Group B (n=404)	Group C (n=405)
At end of trial or when last seen alive					
Remained on first-line ART	578 (95%)	565 (94%)	371 (93%)	387 (96%)	385 (95%)
On randomised first-line ART	534 (88%)	519 (87%)	330 (83%)	356 (88%)	367 (91%)
Had switched to second-line ART	28 (5%)	35 (6%)	26 (7%)	17 (4%)	20 (5%)
For CD4 criteria, per protocol	0	17	10	3	4
For other CD4 concerns*	0	11	3	4	4
For new or recurrent WHO stage 4 event per protocol†	26	6	11	10	11
Severe failure to thrive	15	3	5	7	6
For persistent WHO stage 3 event per protocol	2	1	2	0	1
Moderately severe failure-to-thrive	1	0	0	0	1
Ever changed a drug in first-line ART					
All	69 (11%)	68 (11%)	46 (12%)	53 (13%)	38 (9%)
Ever changed a drug for toxicity‡	40 (7%)	44 (7%)	19 (5%)	35 (9%)	30 (7%)
Ever substituted a drug for toxicity	27 (4%)	26 (4%)	19 (5%)	17 (4%)	17 (4%)
Ever changed a drug for tuberculosis‡	31 (5%)	28 (5%)	30 (8%)	20 (5%)	9 (2%)
Ever substituted a drug for tuberculosis	24 (4%)	24 (4%)	30 (8%)	15 (4%)	3 (1%)
Total changes to first-line ART					
All	75 (100%)	76 (100%)	53 (100%)	58 (100%)	40 (100%)
Total changes for toxicity	29 (39%)	30 (39%)	5 (9%)	28 (48%)	26 (65%)
Dropped zidovudine for haematological AE	11	13	0	15	9
Substituted zidovudine for haematological AE	5	5	0	1	9
Changed for rash or hypersensitivity	6	7	3	6	4
Changed for raised liver function tests	3	2	1	3	1
Changed for lipodystrophy or lipoatrophy	4	3	1	3	3
Total changes for tuberculosis	31 (41%)	28 (37%)	30 (57%)	20 (34%)	9 (23%)
Dropped nevirapine	7	4	0	5	6
Substituted nevirapine with another drug	24	24	30	15	3
Total changes for other reasons§	15 (20%)	18 (24%)	18 (34%)	10 (17%)	5 (13%)

Data are n (% of children). ART=antiretroviral therapy. AE=adverse event. *Typically CD4 slightly higher than the WHO thresholds (confirmed on-ART CD4 <15% if age 1–2 years, <10% if 3–4 years, <100 cells per mL if ≥5 years). †Some events that led to switching were not adjudicated to meet prespecified protocol criteria by the endpoint review committee and so were not included as primary or secondary endpoints. ‡Children randomly assigned to four-drug induction could drop one of the four original drugs for toxicity or other reasons, without having to substitute. §11 dispensing errors, six changes due to pregnancy, 16 other carer preference or domestic reasons.

Table 2: ART received

100 child-years, 95% CI 0·37–1·60, $p=0\cdot002$). The most common WHO stage 4 events were oesophageal candidiasis (eight on clinically driven and five on routine laboratory monitoring) and severe unexplained failure-to-thrive (nine on clinically driven and three on routine laboratory monitoring).

25 (4%) children died in the clinically driven monitoring group versus 29 (5%) in the routine laboratory monitoring group (1·1 vs 1·3 per 100 child-years, respectively; difference $-0\cdot2$ per 100 child-years, 95% CI $-0\cdot82$ to $0\cdot41$; HR 0·84, 95% CI 0·49–1·44; $p=0\cdot45$; appendix). Most deaths (19 clinically driven and 20 routine laboratory monitoring) were primarily HIV-related; only one was drug-related (chemotherapy plus zidovudine). Similar variation over time (heterogeneity $p=0\cdot03$; appendix) was observed for deaths alone as for the combined endpoint of WHO stage 4 or death (13 deaths on clinically driven and 27 on routine laboratory monitoring in first year, 12 vs two subsequently; difference 0·56 per 100 child-years, 95% CI 0·15–0·97, with 12 of the 14 deaths after 1 year in children aged >8 years). Progression to new WHO stage 3 or 4 events or death gave similar results (HR [clinically driven:routine laboratory monitoring] 1·00, 0·73–1·38; $p=0\cdot98$; appendix). Pulmonary tuberculosis was the commonest WHO stage 3 event (25 clinically driven monitoring, 28 routine laboratory monitoring). Among 45 WHO stage 3 or 4 events reported on clinically driven monitoring and 21 on routine laboratory monitoring from the second year onwards, 14 versus one were failure-to-thrive and 15 versus 15 were extrapulmonary or pulmonary tuberculosis, highlighting the potential of weight monitoring to identify first-line CD4 failure clinically.

CD4 percentage increased throughout the first 3 years on ART before plateauing in both groups (figure 3; $p=0\cdot23$). Only 11 (2%) children on clinically driven monitoring versus two ($<1\%$) on routine laboratory monitoring had CD4 less than 5% at their last visit (exact $p=0\cdot01$; appendix). Viral load suppression was similar in both monitoring groups (figure 3; global $p>0\cdot7$). At the latest test, median 3·7 (IQR 3·0–4·1) years after ART initiation, 351 (77%) of 458 children on clinically driven monitoring versus 345 (78%) of 443 on routine laboratory monitoring had viral load less than 400 copies per mL ($p=0\cdot66$), similar across ages (heterogeneity $p=0\cdot25$); 329 (72%) on clinically driven monitoring versus 313 (71%) on routine laboratory monitoring had viral load less than 80 copies per mL ($p=0\cdot70$).

Weight for age and height for age did not differ significantly between groups ($p=0\cdot71$, $p=0\cdot07$; appendix). 49 (9%) children on clinically driven monitoring versus 29 (5%) on routine laboratory monitoring had weight-for-age Z score less than -3 (approximate one thousandth percentile of normal UK weight) at last visit (global $p=0\cdot12$).

One or more grade 3 or 4 adverse events (co-primary endpoint) occurred in 283 (47%) children on clinically

driven monitoring versus 282 (47%) on routine laboratory monitoring (HR 0·98, 95% CI 0·83–1·16, $p=0\cdot83$; figure 3). Of 1170 adverse events (621 clinically driven monitoring, 549 routine laboratory monitoring), 810 (69%) were asymptomatic laboratory results, most commonly grade 3 neutropenia (171 clinically driven monitoring, 167 routine laboratory monitoring; appendix); only 87 (7%) were definitely, probably, or uncertainly ART-related (41 clinically driven monitoring, 46 routine laboratory monitoring). 111 (18%) children on clinically driven monitoring versus 109 (18%) on routine laboratory monitoring had one or more grade 4 adverse events (HR 0·99, 0·76–1·29, $p=0\cdot94$). Although there was no difference in grade 4 adverse events, 147 (24%) children on clinically driven monitoring versus 117 (20%) on routine laboratory monitoring had one or more serious adverse events (any grade; HR 1·30, 1·02–1·66, $p=0\cdot04$). Most of the 362 serious adverse events (217 clinically driven, 145 routine laboratory monitoring) were malaria (113 clinically driven, 65 routine laboratory monitoring), and most (179 clinically driven, 117 routine laboratory

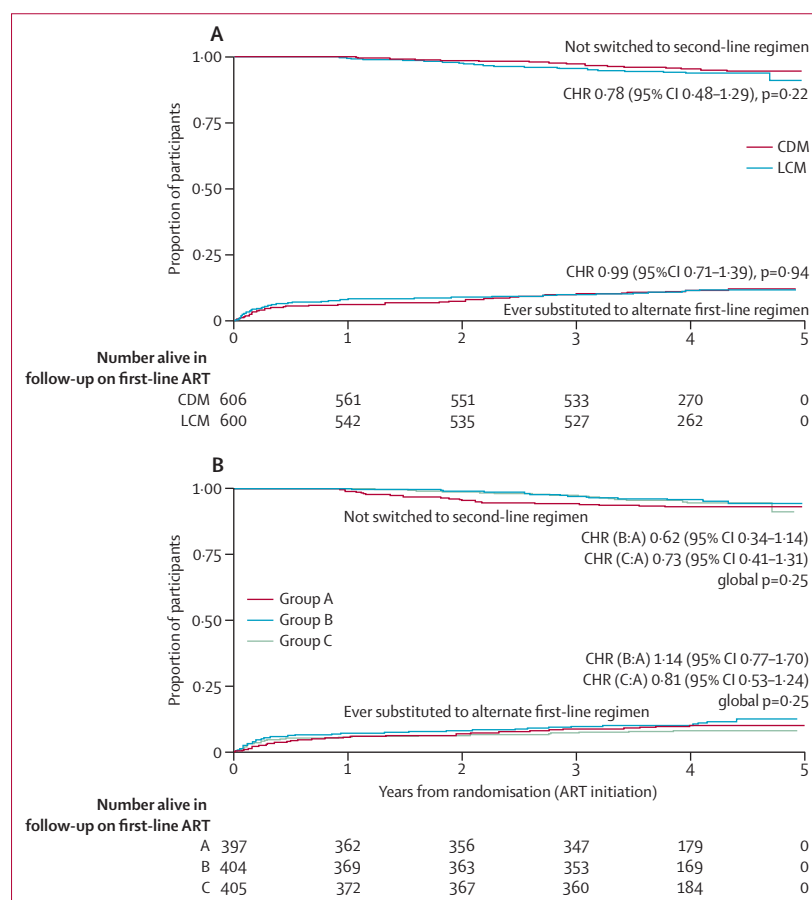
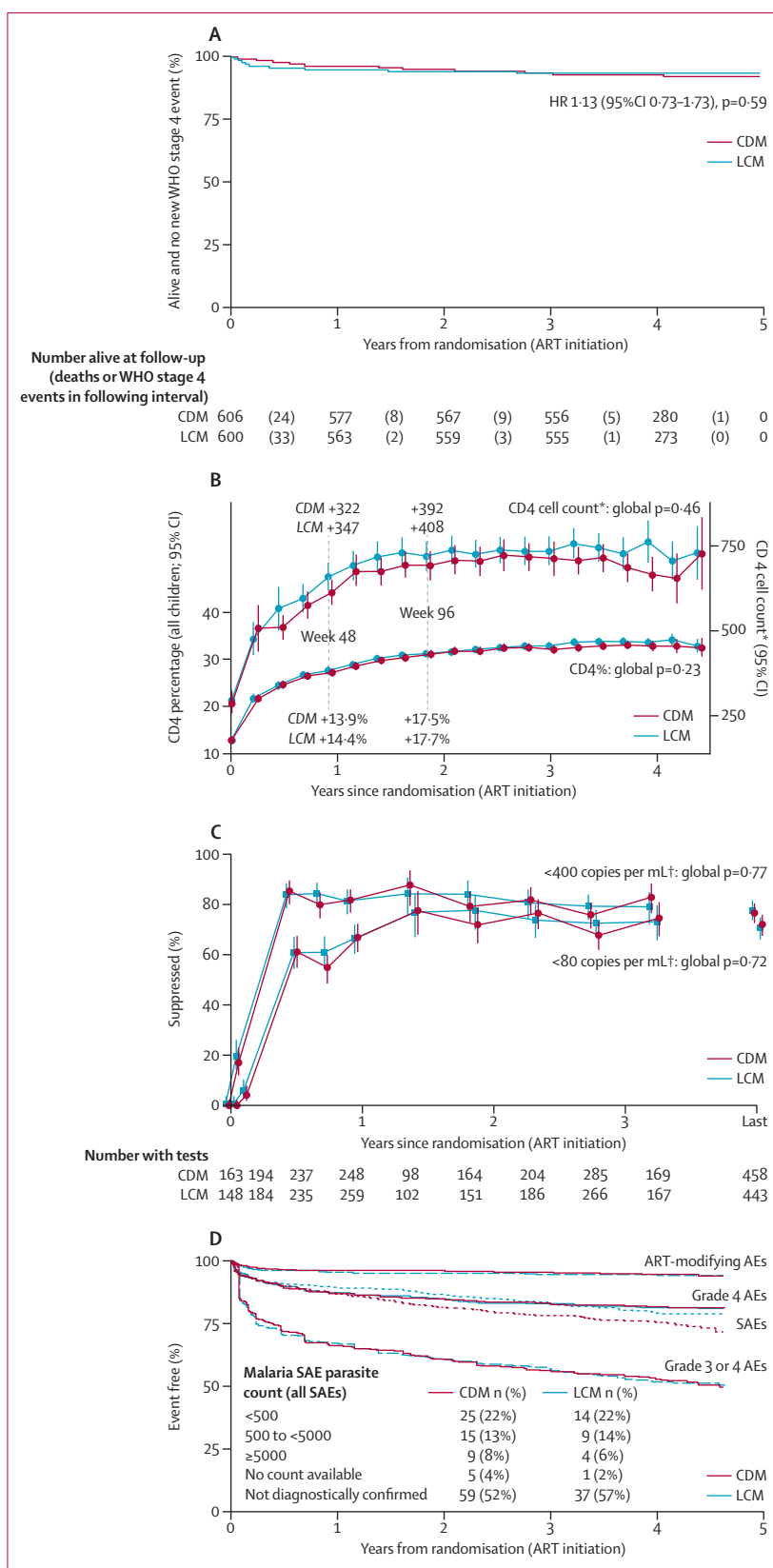


Figure 2: Substitution in first-line ART (any reason) and switch to second-line ART (A) By monitoring strategy. (B) By ART strategy. Lines show cumulative incidence of switch to second-line treatment before death on first-line, and of substitution in first-line before death on first-line estimated treating death on first-line as a competing risk. ART=antiretroviral therapy. CHR=cause-specific hazard ratio. CDM=clinically driven monitoring. LCM=laboratory and clinical monitoring.



monitoring) were admissions to hospital. The excess malaria serious adverse events in the clinically driven monitoring group were mostly in children with parasite counts less than 500 per 200 white blood cells, or were not diagnostically confirmed (figure 3). Differences in time to first hospital admission were smaller (HR 1.18, 0.99–1.41, $p=0.07$), with no difference in duration of admission (median 5 [IQR 3–6] days in clinically driven and routine laboratory monitoring; rank-sum $p=0.54$). ART-modifying adverse events occurred in 31 (5%) children on clinically driven monitoring versus 32 (5%) on routine laboratory monitoring (HR 0.95, 0.58–1.56, $p=0.84$). The most common modification (14 clinically driven, 13 routine laboratory monitoring) was to stop (on four-drug regimen) or substitute zidovudine.

There was no significant difference between the three ART strategy randomisation groups in mean CD4 percentage change at week 72 ($p=0.33$) or 144 ($p=0.69$; figure 4). However, at week 36 (when all children moved to three drugs), CD4 percentage responses were significantly greater in the four-drug induction groups ($p<0.0001$; figure 4; appendix).

Viral load suppression less than 400 copies per mL was similar in the three ART strategy groups at weeks 4, 36, and 48 ($p>0.4$), but differed significantly at weeks 24 ($p=0.009$) and 144 ($p=0.009$; figure 4). At week 24, suppression was significantly greater in induction groups receiving four drugs (285 [88%] of 324 children in groups B and C vs 114 [77%] of 148 in group A). By contrast, similarly to week 144, at the latest test suppression was significantly greater in children receiving two NRTIs plus an NNRTI (496 [84%] of 591 in groups A and B vs 200 [65%] of 310 in group C). No significant difference was seen in suppression less than 80 copies per mL at week 24 ($p=0.41$); however, suppression less than 80 copies per mL was significantly greater in groups A and B at week 144 and latest test (466 [79%] in groups A and B vs 176 [57%] in group C; $p<0.0001$). Results were similar when restricted to children younger than 5 years (not shown) or 3 years (appendix).

There was no evidence of differences between ART-strategy groups in progression to new WHO stage 4 event or death (HR [B:A] 0.89, 95% CI 0.53–1.48; HR [C:A] 0.91, 0.54–1.52; global $p=0.89$), WHO stage 3 or 4 event or death (HR [B:A] 0.82, 0.55–1.20; HR [C:A] 0.80, 0.54–1.17; global $p=0.44$), or mortality overall (HR [B:A] 0.66, 0.33–1.31; HR [C:A] 0.97, 0.52–1.81; global $p=0.43$; figure 4; appendix). In particular, there

Figure 3: New WHO 4 stage event or death (A), CD4 (B), viral load (C), and adverse events (D), by monitoring strategy

CDM=clinically driven monitoring. LCM=laboratory and clinical monitoring. ART=antiretroviral therapy. AE=adverse event. SAE=serious adverse event.

*CD4 cell count in children aged 5 years or older at ART initiation. †Upper lines show suppression less than 400 copies per mL and lower lines suppression less than 80 copies per mL (2:1 dilution used for most samples because of low stored volumes).

was no evidence that greater initial CD4 increases in groups B or C significantly reduced disease progression risks in the first year (13 deaths in group A vs nine in group B vs 18 in group C; 22 vs 19 vs 24 WHO stage 4 events or deaths; 36 vs 35 vs 37 WHO stage 3 or 4 events or deaths). However, there was also no suggestion of higher event rates in group C receiving long-term three-NRTI maintenance, despite lower long-term viral load suppression; if anything, fewer events occurred after 1 year (seven deaths in group A vs five in group B vs two in group C; 18 vs 13 vs eight WHO stage 4 events or deaths; 37 vs 26 vs 17 WHO stage 3 or 4 events or deaths). There was no evidence of differences in weight for age ($p=0.58$) or height for age ($p=0.90$) across ART strategies (appendix).

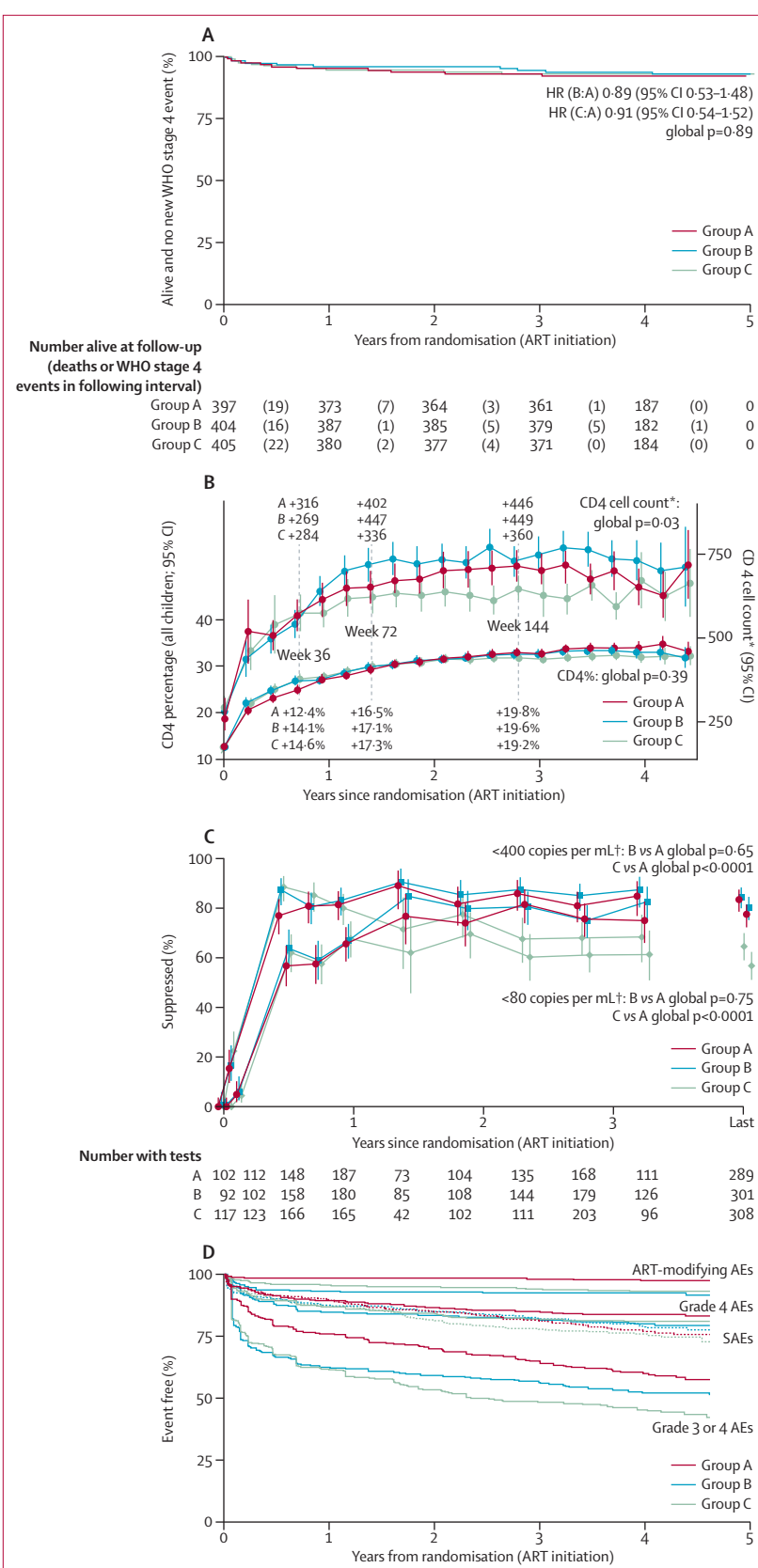
157 (40%) children in group A, 190 (47%) in group B, and 218 (54%) in group C had one or more grade 3 or 4 adverse events (co-primary endpoint; HR [B:A] 1.32, 95% CI 1.07–1.63; HR [C:A] 1.58, 1.29–1.94; global $p=0.0001$; figure 4). The difference was almost exclusively driven by excess asymptomatic neutropenia (86 vs 133 vs 184 events; appendix). There were 15 versus 41 versus 31 definitely, probably, or uncertainly ART-related grade 3 or 4 adverse events, respectively. There were no significant differences in grade 3 or 4 anaemia or grade 4 only anaemia (with or without clinical symptoms; grade 3 or 4: 38 vs 44 vs 44, respectively; grade 4: 21 vs 24 vs 25, respectively). There were also no significant differences in the numbers of children with one or more grade 4 adverse events (63 [16%] vs 81 [20%] vs 76 [19%]; HR [B:A] 1.27, 0.91–1.76; HR [C:A] 1.20, 0.86–1.68; global $p=0.34$) or serious adverse events (87 [22%] vs 82 [20%] vs 95 [23%]; HR [B:A] 0.92, 0.68–1.25; HR [C:A] 1.09, 0.81–1.46; global $p=0.53$).

ART-modifying adverse events occurred in eight (2%) children in group A versus 30 (7%) in group B and 25 (6%) in group C (HR [B:A] 3.80, 95% CI 1.74–8.29; HR [C:A] 3.09, 1.39–6.85; global $p=0.002$; figure 4). The most common modification (13 children in group B, 14 in C) was to stop (on four-drug regimen) or substitute zidovudine because of anaemia, even though grade 3 and 4 anaemias occurred similarly across all three groups. Despite substantial numbers of grade 3 neutropenias, only six children (two in group B, four in group C) modified ART (zidovudine) for this reason. Three children substituted zidovudine because of lipodystrophy in group C; two children in group A and three in group B substituted efavirenz because of lipodystrophy or gynaecomastia.

Figure 4: New WHO 4 stage event or death (A), CD4 (B), viral load (C), and adverse events (D), by ART strategy

ART=antiretroviral therapy. AE=adverse event. SAE=serious adverse event.

*CD4 cell count in children aged 5 years or older at ART initiation. †Upper lines show suppression less than 400 copies per mL and lower lines suppression less than 80 copies per mL (2:1 dilution used for most samples because of low stored volumes).



Discussion

The DART trial⁵ in African adults with HIV showed that routine laboratory monitoring for ART side-effects had no effect on toxicity outcomes. Although routine CD4 monitoring had significant benefits on disease progression and mortality, absolute differences were small.⁵ Health-economic analyses suggested point-of-care CD4 tests would need to cost less than US\$3·80 for monitoring every 12 weeks after the first year on ART to be cost-effective.²⁵ Smaller adult trials of routine CD4 or viral load monitoring, or both, have shown similar results,^{6–8} although no other trial has investigated toxicity monitoring. These results provide strong reassurance that increasing coverage by rollout of ART to adults at lower-level health facilities is the most rational and cost-effective policy at a population level, irrespective of provision of laboratory monitoring.

Most health centres in resource-limited settings treat adults and children with HIV in the same clinics. However, results from studies in adults might not generalise to children who have different comorbidities (eg, more anaemia associated with malnutrition, malaria, and sepsis, which could affect haematological monitoring). Differences in predictive value and interpretation of CD4 tests,^{9,10} faster disease progression in young children,²⁶ and greater sensitivity of ART-related weight gain (which has reflected virological response²⁷) could also affect the relative benefits of laboratory versus clinical monitoring.

We therefore undertook the first paediatric trial investigating ART monitoring in children with fairly advanced HIV disease (panel). We found that NNRTI-based regimens, including WHO-recommended NRTIs, can be delivered safely across childhood without routine toxicity monitoring. Toxicity substitutions were infrequent, as previously reported;³¹ most were because of asymptomatic haematology results in children on zidovudine. Interestingly, grade 3 and 4 and grade 4 anaemias occurred similarly across groups receiving and not receiving zidovudine, both short and long term, suggesting that anaemia in children on ART is most likely caused by chronic HIV infection rather than antiretroviral drugs. Therefore, routine haemoglobin monitoring is likely unnecessary, even in children on zidovudine, and continuation of zidovudine without substitution (as occurred in clinically driven monitoring group C children) is unlikely to result in harm. Further, lack of haemoglobin testing before initiation of ART should not prevent zidovudine use, since haemoglobin values increased after ART initiation. Crucially, we observed no evidence of interaction between monitoring and ART strategies on any toxicity outcome. We previously reported very few possible hypersensitivity reactions to abacavir (four [$<1\%$] in 1206 patients³²) or difficulties managing abacavir plus nevirapine together in African children.

Importantly for programme planners, first-line drug substitutions occurred as often for tuberculosis as for adverse events. Tuberculosis was also by far the most

common WHO stage 3 or 4 event. One advantage of the four-drug induction regimen was the ability to simply drop nevirapine in children starting rifampicin. The ongoing tuberculosis incidence illustrates the potential usefulness of three-NRTI regimens in children, with 91% randomly assigned to group C still receiving maintenance with lamivudine, abacavir, and zidovudine after median 4 years. Tuberculosis remains particularly difficult to manage in children younger than 3 years who cannot take efavirenz, because rifampicin coadministration significantly reduces nevirapine³³ and lopinavir concentrations. Super-boosting with additional ritonavir has been recommended with lopinavir,³⁴ but ritonavir is unpalatable and difficult to dose.¹⁵ Substitution to three NRTIs is frequently used for children on NNRTI plus two NRTIs developing tuberculosis, but there have been concerns about reduced efficacy. Viral load suppression was similar to standard NNRTI-based ART at 48 weeks for children moving to maintenance three-NRTI at 36 weeks. However, viral load suppression was significantly lower at 144 weeks, suggesting that long-term three-NRTI treatment would not be advisable (even after four-drug induction). Importantly, there was no evidence of immunological or clinical harm from roughly 15% lower viral load suppression long-term with three NRTIs; if anything this group had fewer clinical events, similar to previous randomised adult data.³⁵ These findings provide reassurance that a three-NRTI regimen is safe for children on ART when they need antituberculosis co-treatment.

Overall, there was no evidence that clinically driven monitoring was inferior to routine laboratory monitoring in terms of disease progression or mortality. However, we found interactions with time on ART, disease progression or death being somewhat lower in the clinically driven monitoring group in the first year. Chance seems the most likely explanation, because management was similar with no child switching to second-line during the first year in both groups. The only other possible explanation is that receiving CD4, biochemistry, and haematology results was actually harmful for children (eg, resulting in clinicians failing to undertake proper clinical evaluations), which seems implausible. From the second year, our results are qualitatively similar to those of the DART trial, with small but significant clinical event excesses in clinically driven monitoring. However, event rates were substantially lower than in DART, so the excess remained within the non-inferiority margin; also fewer children switched to second-line ART with similar proportions in both groups. Unlike DART, there was no evidence of excess switching with very low CD4 in clinically driven monitoring. Whereas switches in routine laboratory monitoring were predominantly triggered by falling CD4, in clinically driven monitoring most switches were for failure to thrive, which might be a more sensitive indicator of first-line CD4 failure than in adults and could be used where CD4 monitoring is unavailable.

Unnecessary switching at high CD4 did occur, as in DART, but in very few children (six on clinically driven monitoring and none on routine monitoring switched with CD4 greater than 25%). Other paediatric studies in which routine CD4 and viral load monitoring were used have also reported fairly low rates of switching from first-line NNRTI-based ART over 5–6 years (eg, 8% of 2570 Ugandan children;³⁶ 22% in the PENPACT-1 trial³⁰).

As expected and previously reported, most deaths were soon after starting ART in children with lowest pre-ART CD4 or weight for age.³⁷ Irrespective of monitoring strategy, 5-year survival was remarkably high (96%, compared with 88% in DART), emphasising the importance of good clinical care, and availability of continuous ART and concomitant treatments. Loss to follow-up was only 2.7%, providing confidence that results are robust. We also observed no differences between clinically driven and routine laboratory monitoring in CD4 or viral load responses. The only difference between monitoring strategies was in serious adverse events, due to an excess of clinical malaria admissions to hospital in clinically driven monitoring. One limitation is that the trial allocation was of necessity open; lack of knowledge of CD4 in a child on clinically driven monitoring presenting with fever could have influenced decisions about hospital admission, given the plausible differential diagnosis of bacteraemia.³⁸

Data for viral load from the 78% of children with results suggest that initiation of ART with four drugs might significantly improve early viral load suppression, consistent with greater early CD4 responses, particularly in those with very low pre-ART CD4. By design all children moved to three drugs at 36 weeks, in view of the possible toxicity and costs of four-drug regimens in the long term. Since toxicity differences were restricted to asymptomatic laboratory results with no effect on ART management, results hint that a longer-term four-drug, three NRTIs plus one NNRTI regimen might have continued to provide viral load and CD4 benefits, although this might not have translated into clinical benefit. However, viral load suppression with two NRTIs plus an NNRTI was fairly high, with 83% of children receiving an NNRTI with abacavir plus lamivudine throughout achieving less than 400 copies per mL long term, irrespective of monitoring strategy, so further gains from a fourth drug might be less plausible.

Viral load suppression was similar to other trials in which viral load monitoring was done routinely (eg, 82% <400 copies per mL after median 5 years on two NRTIs plus an NNRTI or a protease inhibitor;³⁰ 85% of infants <400 copies per mL after median 5 years on two NRTI plus lopinavir or ritonavir (A Violari, personal communication); 85% and 75% of children <3 years <400 copies per mL after 48 weeks on two NRTI plus lopinavir or ritonavir, or nevirapine, respectively²⁸). This finding might be partly because of superiority of abacavir (in the ARROW two-NRTI backbone) over zidovudine, as

Panel: Research in context

Systematic review

We searched PubMed up to Jan 5, 2013, with search terms “HIV” AND “monitoring” AND (“antiretroviral therapy” OR “ART”) AND “trial”*. We identified three trials in adults;^{5,7,8} one further trial had been presented but not published.⁶ There were no trials in children. Trials in adults showed that clinical monitoring was safe and feasible, but CD4 monitoring to detect first-line failure provided small additional benefits; addition of viral load to routine CD4 monitoring provided no further benefits. Replacement of “monitoring” with child* identified previous short-term (24 or 48 week) trials that showed three-drug protease inhibitor-based (lopinavir or ritonavir) ART to be superior to nevirapine-based ART in children younger than 3 years;^{28,29} however, studies in older children found similar responses,³⁰ and lopinavir and ritonavir are costly and logistically challenging to administer.

Interpretation

Routine laboratory monitoring for toxicity on non-nucleoside reverse transcriptase inhibitors (NNRTIs) plus nucleoside reverse transcriptase inhibitors (NRTIs; abacavir or zidovudine plus lamivudine) is not needed in children, as in adults; requiring such monitoring might be a barrier to life-saving treatment. CD4 monitoring provided a small but significant reduction in disease progression or death after the second year on ART in children, as in adults. However, unlike adults, CD4 and viral load responses were very similar irrespective of CD4 versus clinical monitoring. Monitoring weight-gain appeared a sensitive indicator of first-line CD4 failure, and drug changes occurred as often for concurrent tuberculosis as for adverse events. Four-drug ART with an NNRTI plus three NRTIs provided superior short-term virological suppression and CD4 responses, but these benefits were not sustained during maintenance three-drug ART. However, 83% of children in ARROW had viral load less than 400 copies per mL with abacavir, lamivudine, and a NNRTI for 3.7 years, irrespective of age, supporting previously reported superiority of abacavir over zidovudine.²⁹ Three-NRTI long-term maintenance was clinically and immunologically similar to NNRTI-based ART, and would be useful during tuberculosis cotreatment. Results support an integrated approach to treatment of adults and children in treatment rollout. Cost and feasibility (eg, through point-of-care tests) of provision of CD4 monitoring are future challenges.

previously reported.²⁷ Coupled with the low clinical event rates after 1 year in ARROW, our data suggest any additional benefits from routine viral load monitoring are likely to be small. Although CD4 and viral load failure do not correlate well,³⁹ the implications of late detection of viral load failure are likely to depend largely on how resistance evolves with persisting viral replication. In PENPACT-1, lamivudine and NNRTI resistance occurred with low-level viral load failure,³⁰ so the consequences of delaying switch on these regimens until CD4 failure might be small; the 3-year PHPT trial in adults, which randomly assigned patients to routine CD4 with or without viral load monitoring, found no additional benefit from viral load, although patients in both groups of this trial had high viral load suppression.⁶

Although open allocation was an unavoidable limitation of the monitoring randomisation and was not undertaken for the ART-strategy randomisation, the endpoint review committee adjudicated endpoints masked to randomisation. In a survey at ARROW exit, only four (1%) of 561 participants on clinically driven monitoring reported having CD4 testing done privately; clinicians remained masked. One possible criticism is that all ARROW centres

had laboratories; however, the only way our results would not generalise to centres with lower-quality clinical care would be if these health-care workers were able to act more appropriately on routine laboratory results than in centres with high-quality care. This scenario seems unlikely since substantial CD4 variability and complexity around toxicity test interpretation mean that simple rules for acting on routine test results are unlikely ever to be optimum. Rather, clinicians providing the best clinical care are plausibly also best able to interpret and act on routine laboratory results. Thus although the overall risks of WHO events or death might be higher under poorer clinical care, differences in outcomes between routine and clinically driven laboratory monitoring would be likely, if anything, to be even smaller than observed in ARROW. Of note, we found no evidence that small benefits from CD4 monitoring varied by pre-ART CD4 or percentage, suggesting our results are robust to changes in ART initiation thresholds.

Toxicity monitoring has no benefit and costs money; it cannot therefore be cost-effective. This fact should reinforce WHO guidelines that routine toxicity tests are not required for paediatric ART provision, as for adults. Formal cost-effectiveness analysis is ongoing: given reduced efficacy of CD4 monitoring compared with DART, CD4 monitoring every 12 weeks is unlikely to be cost-effective in children overall (P Revill, personal communication), although it might have some potential to pick up earlier failure in older children or adolescents concealing adherence challenges. Cost-effectiveness analysis and systematic review is also planned to compare the induction-maintenance ART strategy with NNRTI and protease inhibitor-based first-line ART, including sensitivity analyses to account for increased efficacy but greater cost of abacavir versus zidovudine.

In conclusion, ARROW results should send a strong message to African ART programmes to accelerate ART rollout to children since this process currently lags woefully behind adults.³ The key finding is that ART provides enormous benefits to children and can be delivered safely with good-quality clinical care and without routine toxicity monitoring. For children initiating ART with severe immune suppression, addition of a fourth drug improves short-term immunological and virological responses; whether continuing a four-drug regimen longer-term would be advantageous remains unclear. However, ARROW results support short-term use of three-drug NRTI regimens during antituberculosis treatment in children already on ART. Simple point-of-care CD4 tests might have a future role, at least to confirm clinical need to switch to second-line ART, as in DART.⁴⁰ In children, monitoring of weight gain should be emphasised as an important additional clinical aid for identification of first-line failure. Laboratory tests remain important for assessment of ART eligibility and for diagnosis and management of intercurrent infections. Mentoring of health-care workers to foster quality clinical care and reassurance that

children do very well without laboratory monitoring of ART should energise faster and further scale-up of ART rollout for children with HIV in Africa.

Contributors

The ARROW trial was designed by DMG, ASW, AK, KN, PMug, SB-K, MT, MB-D, VM, PMun, and PMus. The trial was done in Uganda by AK, PMug, PN-N, SB-K, VM, PMun, PMus, RK, and CT, and in Zimbabwe by KN, MB-D, TV, and AJP; and coordinated in the UK by DMG, MT, BN-J, JC, and ASW. ASW and ADC wrote the trial analysis plan, which all authors then reviewed; and ADC did the analyses. All authors contributed to interpretation of the data. DMG, ADC, and ASW wrote the first draft of the report. All authors revised the report critically and approved the final version.

Conflicts of interest

We declare that we have no conflicts of interest.

Acknowledgments

Analysis and writing committee: Adeodata Kekitiinwa, Adrian Cook, Kusum Nathoo, Peter Mugenyi, Patricia Nahirya-Ntege, Sabrina Bakeera-Kitaka, Margaret Thomason, Mutsa Bwakura-Dangarembizi, Victor Musiime, Paula Munderi, Bethany Naidoo-James, Tichaona Vhembo, Constance Tumusiime, Richard Katuramu, Jane Crawley, Andrew J Prendergast, Philippa Musoke, A Sarah Walker, Diana M Gibb. We thank the children, carers, and staff from all the centres participating in the ARROW trial, and the ARROW trial steering committee for access to data.
MRC/UVRI Uganda Research Unit on AIDS, Entebbe, Uganda: P Munderi, P Nahirya-Ntege, R Katuramu, J Lutaakome, F Nankya, G Nabulime, I Sekamatte, J Kyarimpa, A Ruberantwari, R Sebukyu, G Tushabe, D Wangi, M Musinguzi, M Aber, L Matama, D Nakitto-Kesi.
Joint Clinical Research Centre, Kampala, Uganda: P Mugenyi, V Musiime, R Keishanyu, VD Afayo, J Bwomezi, J Byaruhanga, P Erimu, C Karungi, H Kizito, W S Namala, J Namusanje, R Nandugwa, T K Najjuko, E Natukunda, M Ndigendawani, S O Nsiyona, R Kibenge, B Bainomuhwezi, D Sseremba, J Tezikyabiri, C S Tumusiime, A Balaba, M Mugumya, F Nghania, D Mwebesa, M Mutumba, E Bagurukira, F Odongo, S Mubokya, M Ssenyonga, M Kasango, E Lutalo, P Ornon.
University of Zimbabwe, Harare, Zimbabwe: K J Nathoo, M F Bwakura-Dangarembizi, F Mapinge, E Chidziva, T Mhute, T Vhembo, R Mandidewa, M Chipiti, R Dzapasi, C Katanda, D Nyoni, G C Tinago, J Bhiri, S Mudzingwa, D Muchabaiwa, M Phiri, V Masore, C C Marozva, S J Maturure, S Tsikirayi, L Munetsi, K M Rashirai, J Steamer, R Nhema, W Bikwa, B Tambawoga, E Mufuka.
Baylor College of Medicine Children's Foundation Uganda, Mulago Hospital, Uganda: A Kekitiinwa, P Musoke, S Bakeera-Kitaka, R Namuddu, P Kasirye, A Babirye, J Asello, S Nakalanzi, N C Ssemambo, J Nakafeero, J Tikabibamu, G Musoba, J Ssanyu, M Kisekka.
MRC Clinical Trials Unit, London, UK: D M Gibb, M J Thomason, A S Walker, A D Cook, B Naidoo-James, M J Spyer, C Male, A J Glabay, L K Kendall, J Crawley, A J Prendergast.
Independent ARROW trial monitors: I Machingura, S Ssenyonjo.
Trial steering committee: I Weller (chair), E Luyirika, H Lyall, E Malianga, C Mwansambo, M Nyathi, F Miiro, D M Gibb, A Kekitiinwa, P Mugenyi, P Munderi, K J Nathoo, A S Walker; *Observers* S Kinn, M McNeil, M Roberts, W Snowden.
Data and safety monitoring committee: A Breckenridge (chair), A Pozniak, C Hill, J Matenga, J Tumwine.
Endpoint review committee (independent members): G Tudor-Williams (Chair), H Barigye, H A Mujuru, G Ndeezi; *Observers* S Bakeera-Kitaka, M F Bwakura-Dangarembizi, J Crawley, V Musiime, P Nahirya-Ntege, A Prendergast, M Spyer.
Economics group: P Revill, T Mabugu, F Mirimo, S Walker, M J Sculpher.
Funding: ARROW is funded by the UK Medical Research Council and the UK Department for International Development. ViiV Healthcare/GlaxoSmithKline donated first-line drugs for ARROW and provided funding for viral load assays.

References

- 1 Gilks CF, Crowley S, Ekpini R, et al. The WHO public-health approach to antiretroviral treatment against HIV in resource-limited settings. *Lancet* 2006; **368**: 505–10.

- 2 Institute for Health Metrics and Evaluation. Financing global health 2011: continued growth as MDG deadline approaches. <http://www.healthmetricsandevaluation.org/publications/policy-report/financing-global-health-2011-continued-growth-mdg-deadline-approaches> (accessed Oct 25, 2012).
- 3 UNAIDS. World AIDS Day report 2012. Geneva: UNAIDS, 2012.
- 4 Abongomera G, Namata H, Nkhata M, et al. Lablite: baseline mapping survey of decentralised ART service provision in Malawi, Uganda and Zimbabwe. XIX International AIDS Conference; Washington DC, USA; July 22–27, 2012; abstr LBPE572012.
- 5 DART Trial Team. Routine versus clinically driven laboratory monitoring of HIV antiretroviral therapy in Africa (DART): a randomised non-inferiority trial. *Lancet* 2010; **375**: 123–31.
- 6 Jourdain G, Ngo-Giang-Huong N, Le Coeur S, et al. PHPT-3: a randomized clinical trial comparing CD4 vs viral load ART monitoring/switching strategies in Thailand. 18th Conference on Retroviruses and Opportunistic Infections. Boston, MA, USA; Feb 27–March 2, 2011; abstr 44.
- 7 Laurent C, Kouanfack C, Laborde-Balen G, et al. Monitoring of HIV viral loads, CD4 cell counts, and clinical assessments versus clinical monitoring alone for antiretroviral therapy in rural district hospitals in Cameroon (Stratall ANRS 12110/ESTHER): a randomised non-inferiority trial. *Lancet Infect Dis* 2011; **11**: 825–33.
- 8 Mermin J, Ekwaru JP, Were W, et al. Utility of routine viral load, CD4 cell count, and clinical monitoring among adults with HIV receiving antiretroviral therapy in Uganda: randomised trial. *BMJ* 2011; **343**: d6792.
- 9 HIV Paediatric Prognostic Markers Collaborative Study Group (HPPMCS). Short-term risk of disease progression in HIV-1-infected children receiving no antiretroviral therapy or zidovudine monotherapy: a meta-analysis. *Lancet* 2003; **362**: 1605–11.
- 10 HIV Paediatric Prognostic Markers Collaborative Study Group (HPPMCS). Predictive value of absolute CD4 cell count for disease progression in untreated HIV-1-infected children. *AIDS* 2006; **20**: 1289–94.
- 11 Cross Continents Collaboration for Kids (3Cs4kids). Markers for predicting mortality in untreated HIV-infected children in resource-limited settings: a meta-analysis. *AIDS* 2008; **22**: 97–105.
- 12 Bartlett JA, Fath MJ, Demasi R, et al. An updated systematic overview of triple combination therapy in antiretroviral-naïve HIV-infected adults. *AIDS* 2006; **20**: 2051–64.
- 13 van Rossum AM, Fraaij PL, de Groot R. Efficacy of highly active antiretroviral therapy in HIV-1 infected children. *Lancet Infect Dis* 2002; **2**: 93–102.
- 14 Judd A. Early antiretroviral therapy in HIV-1-infected infants, 1996–2008: treatment response and duration of first-line regimens. *AIDS* 2011; **25**: 2279–87.
- 15 WHO. Antiretroviral therapy for HIV infection in infants and children: towards universal access. Recommendations for a public health approach. Geneva: World Health Organization, 2006.
- 16 Ssali F, Stohr W, Munderi P, et al. Prevalence, incidence and predictors of severe anaemia with zidovudine-containing regimens in African adults with HIV infection within the DART trial. *Antivir Ther* 2006; **11**: 741–49.
- 17 National Institutes of Health Division of AIDS. Division of AIDS table for grading the severity of adverse events. Bethesda, MD: National Institute of Allergy and Infectious Diseases, 2004 (clarification 2009).
- 18 National Institutes of Health Division of Microbiology and Infectious Diseases (DMID) pediatric toxicity tables. <http://www.niaid.nih.gov/LabsAndResources/resources/DMIDClinRsrch/Documents/dmidpedtox.pdf> (accessed Feb 11, 2013).
- 19 Kasiyire P, Kendall L, Adkison KK, et al. Pharmacokinetics of antiretroviral drug varies with formulation in the target population of children with HIV-1. *Clin Pharmacol Ther* 2012; **91**: 272–80.
- 20 Musiime V, Kendall L, Bakeera-Kitaka S, et al. Pharmacokinetics and acceptability of once- versus twice-daily lamivudine and abacavir in HIV type-1-infected Ugandan children in the ARROW Trial. *Antivir Ther* 2010; **15**: 1115–24.
- 21 WHO. Annex E: prescribing information and weight-based dosing of available ARV formulations for infants and children. Geneva: World Health Organization, 2010.
- 22 International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. Clinical safety data management: definitions and standards for expedited reporting (E2A). London: European Medicines Agency, 1994.
- 23 Paediatric European Network for the Treatment of AIDS. Comparison of dual nucleoside-analogue reverse-transcriptase inhibitor regimens with and without nevirapine in children with HIV-1 who have not previously been treated: the PENTA 5 randomised trial. *Lancet* 2002; **359**: 733–40.
- 24 Wade AM, Ades AE. Age-related reference ranges: significance tests for models and confidence intervals for centiles. *Stat Med* 1994; **13**: 2359–67.
- 25 Medina Lara A, Kigozi J, Amurwon J, et al. Cost effectiveness analysis of clinically driven versus routine laboratory monitoring of antiretroviral therapy in Uganda and Zimbabwe. *PloS One* 2012; **7**: e33672.
- 26 Violari A, Cotton M, Otwombe K, et al. Does early initiation of ART in infants affect virologic and resistance outcomes? Data from the CHER trial after 6 years of follow up. HIV11 Congress; Glasgow, UK; Nov 11–14; abstr 0224.
- 27 Green H, Gibb DM, Walker AS, et al. Lamivudine/abacavir maintains virological superiority over zidovudine/lamivudine and zidovudine/abacavir beyond 5 years in children. *AIDS* 2007; **21**: 947–55.
- 28 Violari A, Lindsey JC, Hughes MD, et al. Nevirapine versus ritonavir-boosted lopinavir for HIV-infected children. *N Engl J Med* 2012; **366**: 2380–89.
- 29 Prendergast AJ, Penazzato M, Cotton M, et al. Treatment of young children with HIV infection: using evidence to inform policymakers. *PLoS Med* 2012; **9**: e1001273.
- 30 The PENPACT-1 (PENTA 9/PACTG 390) Study Team. First-line antiretroviral therapy with a protease inhibitor versus non-nucleoside reverse transcriptase inhibitor and switch at higher versus low viral load in HIV-infected children: an open-label, randomised phase 2/3 trial. *Lancet Infect Dis* 2011; **11**: 273–83.
- 31 Mulenga V, Cook A, Walker AS, et al. Strategies for nevirapine initiation in HIV-infected children taking pediatric fixed-dose combination “baby pills” in Zambia: a randomized controlled trial. *Clin Infect Dis* 2011; **51**: 1081–89.
- 32 Nahirya-Ntege P, Musiime V, Naidoo B, et al. Low incidence of abacavir hypersensitivity reaction among African children initiating antiretroviral therapy. *Pediatr Infect Dis J* 2010; **30**: 535–37.
- 33 Oudijk JM, McIlleron H, Mulenga V, et al. Pharmacokinetics of nevirapine in HIV-infected children under 3 years on rifampicin-based antituberculosis treatment. *AIDS* 2012; **26**: 1523–28.
- 34 Ren Y, Nuttall JJ, Egbers C, et al. Effect of rifampicin on lopinavir pharmacokinetics in HIV-infected children with tuberculosis. *J Acquir Immune Defic Syndr* 2008; **47**: 566–69.
- 35 Munderi P, Walker AS, Kityo C, et al. Nevirapine/zidovudine/lamivudine has superior immunological and virological responses not reflected in clinical outcomes in a 48-week randomized comparison with abacavir/zidovudine/lamivudine in HIV-infected Ugandan adults with low CD4 cell counts. *HIV Med* 2010; **11**: 334–44.
- 36 Musiime V, Kaudha E, Kayiwa J, et al. Antiretroviral drug resistance profiles and response to second-line therapy among HIV-1 infected Ugandan children. *AIDS Res Hum Retroviruses* 2013; published online Jan 11. DOI:10.1089/aid.2012.0283.
- 37 Walker AS, Prendergast AJ, Mugenyi P, et al. Mortality in the year following antiretroviral therapy initiation in HIV-infected adults and children in Uganda and Zimbabwe. *Clin Infect Dis* 2012; **55**: 1707–18.
- 38 Berkley JA, Bejon P, Mwangi T, et al. HIV infection, malnutrition, and invasive bacterial infection among children with severe malaria. *Clin Infect Dis* 2009; **49**: 336–43.
- 39 Walker AS, Gibb DM. Monitoring of highly active antiretroviral therapy in HIV infection. *Curr Opin Infect Dis* 2011; **24**: 27–33.
- 40 Gilks CF, Walker AS, Munderi P, et al. A single CD4 test with 250 cells/mm³ threshold predicts viral suppression in HIV-infected adults failing first-line therapy by clinical criteria. *PloS One* (in press).